# 项目报告

**时间**： 2024 年 4 月 24 日

**团队成员**： 郝馳宇、常烁晨、刘浩源、杜昱泽

**使用的分析工具或开发工具**：D3，Echarts，Vue

**1. 数据处理**

在本项目的数据处理过程中，我们发现所需的json文件可以分为两类：一类为“01-2023年全年图书借阅预约归还等数据.json”、“02-2023年全年入馆数据.json”等数据，这些文件包含数据量大，读取时间较长；另一类为“00-图书馆资源总量.json”等数据，这些文件内容较为简单。

对此，我们对第一类文件按照日期进行切分，这样程序在运行时只需要读取一定范围内的文件，避免过高时延，同时支持了时序分析。对于第二类文件，我们将数据直接硬编码存储中，加快运行速度。

在数据处理过程中，我们发现数据有一定瑕疵，如入馆记录中的身份信息字符串在文字后尾随数个空格，可能导致图表无法正常显示文字，对此我们在数据处理过程中对该字段执行trim操作，删去文字尾随的空格。此外，部分记录存在数据缺失，格式不统一（正常数据/null /“N/A”）的问题，对此，我们在处理数据时保留了缺失但是对可视化无影响的记录，对于非法格式则全部过滤。

数据读取过程中，数据被存储在Map结构中，按照模块功能实现所需的字段作为主键进行存储，如在借阅信息json的读取过程中，可以将入馆身份信息作为主键存储其在记录中的出现次数以支持借阅身份比例显示。与此同时，另一个Map结构将借阅书籍名称作为主键，存储该书籍的借阅次数以支持借阅排行功能。

**2. 视图设计**



图 1 图书馆数据可视化总览

本项目的可视化网页结果如图 1所示，其中可以分为标题栏，左，右上，右中，右下五个部分。以下按该顺序对各视图进行介绍。

标题栏位于页面最上，“图书馆数据可视化”展示该页面名称，左侧的日期选单在显示当前选定时间的同时可以进一步设置，以此改变部分模块的展示数据。

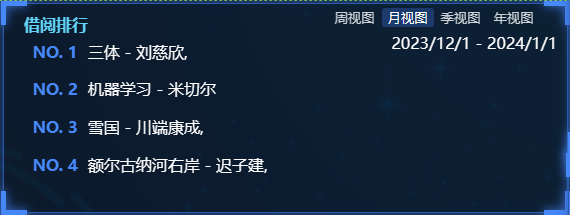


图 2 借阅排行视图时间跨度切换示意

左侧部分分为四个子视图，主要展示图书馆入馆与书籍借阅相关数据。最上方的读者画像视图，展示了2023年全年的入馆人次与借阅人次，旁边采用环形图展示2023年度入馆人员的身份分布比例。中间的读者学院分布视图使用条形图，展示了2023年度入馆人次前五名的学院，并在此基础上展示了对应学院的书籍借阅次数。下方的借阅排行视图则与日期选单联动，当日期更改时，借阅排行自动更新到对应时间跨度内被借阅次数最多的书籍排行，使用文字显示具体时间跨度。该视图提供周/月/季/年四个按钮，使得读者能在不同时间跨度长度内切换，如图 2所示。最下方的图书馆资源总览视图展示了图书馆截至目前的资源存储情况，如当前纸质馆藏册数等。

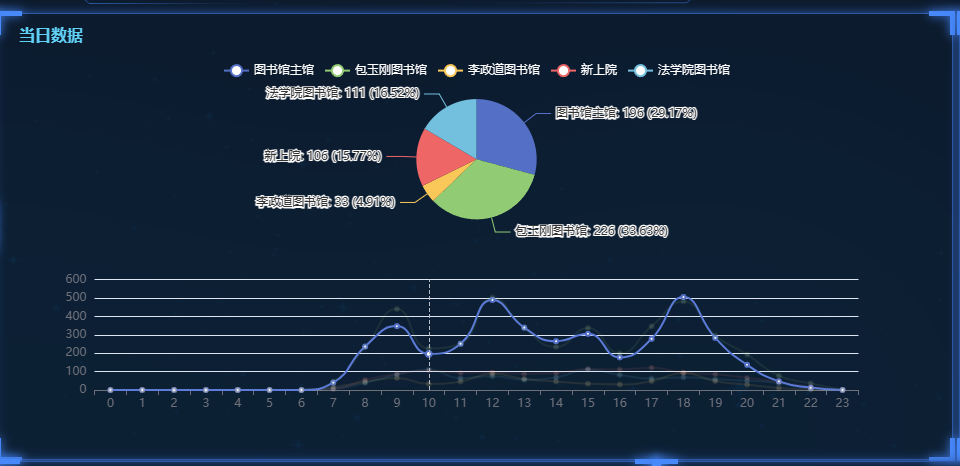


图 3 当日数据视图示意

右上方的当日数据视图是该项目最大的视图。该视图由两部分组成，且与日期选单联动，下方使用折线图展示当日24小时内上海交通大学图书馆的在馆人数随时间的变化情况，将鼠标放在任意一条折线上时，一方面该折线会高亮显示，另一方面上方将会出现扇形图展示该时刻四个图书馆的准确在馆人数及比例，如图 3所示。此外，新书榜单视图使用提供的新书榜单数据，使用金字塔图展示前三本入库新书的书名与作者信息

右中的图书借还趋势视图使用折线图表示选定日期当月图书馆的书籍借阅/归还/续借/预约情况，并使用白色文字直接在左侧显示具体数值。检索热度视图则显示全年书籍检索中的高热度关键词。

屏幕上有字

描述已自动生成

图 4 共享办公预约视图示意

电脑显示屏

中度可信度描述已自动生成

图 5 空间预约视图示意

右下包含两个视图，主要展示图书馆共享空间的占用信息。共享办公预约视图提供了图书馆提供的共享办公位置预约信息，并提供选单在不同位置信息间进行切换。该视图使用折线柱状图展示某位置2023全年的预约信息，其中，折线图展示预约人数随月份变化的信息，而柱状图则展示实际预约人次；此外，我们仿照热力图，使用柱形的颜色深浅来表示预约时长，如图 4所示。空间预约视图则展示图书馆小组讨论室/面试空间的占用信息。其中横轴表示一天24小时时间跨度，纵轴则从日期选单开始向后推进七天，如果选定的共享空间在某一时段为无人使用状态，该时段将以灰色表示，若有人使用则以绿色表示。该视图提供两个选单，第一个选单在图书馆之间切换，第二个选单自动刷新并展示对应图书馆所有共享空间供选择，如图 5所示。

1. **分析流程与结论**

通过分析读者画像子视图，我们可以发现全年入馆人员本科生/硕士/博士/教师比例大致相同，考虑到校内身份实际比例，我们可以发现教师入馆频率大于学生。通过分析读者学院分布视图，我们能够发现电子信息与电气工程学院入馆人次远超其他学院，与学院的人员比例不协调，据此我们可以推断电子信息与电气工程学院学业压力较大，导致该学院学生进入图书馆自习的频率更高。此外，借阅人数前四名大致与入馆人数成正比，而船舶海洋与建筑工程学院人员的借阅率明显高于其余学院，达到10%以上，我们认为出现这种现象的原因是相较其他学院，船院专业书籍出版量更少，价格较高，导致该学院人员更多选择图书馆借阅而非自费购买。通过分析借阅排行视图，一方面小说/通俗历史读物等类型书籍排行靠前，说明同学更倾向于借阅休闲书籍；另一方面，我们不难发现春秋学期考试周前教辅书籍的借阅量明显上升。

通过分析每日数据视图，我们发现闵行校区主图书馆与包玉刚图书馆在馆人数基本保持一致，而徐汇校区新上院人数较少而大于李政道图书馆，这是因为李政道图书馆内座位较少，无法容纳太多人员。此外，两个人数较多的图书馆在馆人数在每日9时、12时、15时，18时存在明显高峰，结合上海交通大学时间表不难发现，9时前绝大多数学生起床并前往图书馆学习，出现第一个高峰，同时也与随后的12、15、18时同时为学生主要下课时间，学生在下课后往往前往图书馆继续学习，因此这些时段在馆人数明显增加。通过分析空间预约视图，我们能够发现预约人员倾向于在下午与晚上使用共享空间，说明小组讨论与面试主要在下午与晚上进行。

通过分析检索热度，人工智能、自动化、深度学习、大数据等关键字占据多个高热度检索词，说明人工智能技术已经成为当前的新趋势被大家普遍学习。此外，通过分析借阅趋势视图，能够发现图书馆在学期结束时出现借书潮，在学期开始时出现还书潮。

1. **组内分工**

本项目中，左上三个视图由郝馳宇完成，左下的图书馆资源总览与右上视图由刘浩源完成，右中视图由常烁晨完成， 右下视图由杜昱泽完成。其中每一视图所需要的数据由该视图的实现者自行处理。项目最后的CSS美化、比例/字体/颜色调整与共享空间视图热力功能由郝馳宇完成。

本报告由郝馳宇撰写。

1. **总结**

本项目全面展示了上海交通大学图书馆相关数据，通过易于理解的页面和合理的模块组织使得用户能快速获取相关信息，并通过日期选择、跨度选择、房间选择等组件给予该可视化视图充分的功能性。在此基础上，本项目考虑到了页面的美观程度，进行了一定程度的美化。

本项目的缺点在于，第一，本项目使用Vue模板进行开发，使得每个HTML元素难以直接进行自定义设置，只能通过在JavaScript脚本中通过修改DOM元素进行设置，对模块的配置和增删带来一定不便；第二：本项目迫于时间限制，对于一定数目的元素使用硬编码设置尺寸，只能适配特定比例的显示器，在其他比例下显示时会出现一定错位、文本或图像较难辨识的问题。

本项目在未来还能进一步改进的部分包括，第一，使用自定义程度更高的框架进行可视化工作，提升修改/增删组件的自由度；第二，减少硬编码的数目，使得网页展示能够根据显示比例动态调整，避免模块错位。